



TITLE:

# 情報量規準 IC (Information Criterion) による統計モデルの評価と選択 (I): 気象要素の時系列構造と自己回帰モデルによる精度予測について

AUTHOR(S):

芝, 正己; 佐々木, 功

---

CITATION:

芝, 正己 ...[et al]. 情報量規準 IC (Information Criterion) による統計モデルの評価と選択 (I): 気象要素の時系列構造と自己回帰モデルによる精度予測について. 京都大学農学部演習林報告 1984, 56: 98-120

ISSUE DATE:

1984-11-30

URL:

<http://hdl.handle.net/2433/191803>

RIGHT:

# 情報量規準 IC (Information Criterion) による 統計モデルの評価と選択 (I)

気象要素の時系列構造と自己回帰モデルによる精度予測について

芝 正 己・佐々木 功

Statistical model estimation and identification  
by Information Criterion (I)

Structure of meteorological time series and accuracy of prediction  
through autoregressive model fitting

Masami SHIBA and Isao SASAKI

## 要 旨

自己回帰モデルを導入して気象要素の時系列構造を解明し、情報量規準 IC がモデル構成の過程における与えられた評価方式下で、いかなるふるまいをするかを実証的に推論することが本論の目的である。これらの結果を要約すると：

【1】 日平均気温の時系列構造は、夏季と冬季では明らかに異なり、夏季においてはある期間の経日的傾向が持続されるが、冬季ではほとんどその傾向は認められず、コログラムは単純マルコフ型で示される。

【2】 降雪季の最大積雪深の予測モデルとして5次の AR モデルが選択され、乱数を加えてシミュレーションした擬似系列は、現実のデータ特性とよく似た性質を示す。

最大積雪深の予測モデル

$$x_n = 0.1034x_{n-1} - 0.1769x_{n-2} - 0.1779x_{n-3} + 0.0689x_{n-4} - 0.4108x_{n-5} + \varepsilon_n$$

$$\varepsilon_n \approx N(0, 1062.58)$$

【3】 月雨量、月平均気温の時系列には、年周期あるいは季節変化以外にはいかなる周期性も認められない。季節変動の型を取り去った時系列は、気温系列においてはわずかな持続性（時定数換算35～40ヶ月）が認められるが、雨量系列についてはほとんど持続性がない。従って、雨量は第1近似としては一種の乱数列であると考えられる。月雨量の時系列は、月別にそれぞれ固有の平均値と分散を持つ12の母集団を考え、それから順次任意標本を抽出して作った標本列であると考えて一応大過ないようである。

月雨量の予測モデル

$$x_n = 0.24496x_{n-1} + 0.11744x_{n-2} + \varepsilon_n$$

$$\varepsilon_n \approx N(0, 5136.67)$$

## 月平均気温の予測モデル

$$x_n = 0.41595x_{n-1} + 0.04759x_{n-2} - \dots + 0.06476x_{n-39} - 0.20648x_{n-40} + \varepsilon_n$$

$$\varepsilon_n \sim N(0, 0.90742)$$

【4】一変量時系列の AR モデル選択における IC の評価の違いは、IC の 2 項目の統計量の差として捉えることができる。データが少ない場合は、IC はいずれも同一モデルを選択することが期待されるが、データ数が増加すると BIC は高次のモデルの分解能が悪くなる。CAI, AIC は高次でのモデルに対しても敏感に反応し、データ数  $n$ ・自由パラメータ数  $m$  に対してその適用領域が、AIC では  $2\sqrt{n}$ , CAI では 2 項目の漸近線  $-\frac{2n(n-1)}{m-n+2}$  として規定される。従って、両者は同一モデルを選択する傾向が高まる。高次項のモデルの分解能という観点に立てば、

$$\text{BIC} \ll \text{CAI} < \text{AIC}$$

という関係が成り立ち、AR モデルの評価規準として、BIC に較べて AIC, CAI はよりフレキシブルな情報量であると見做せる。

## KEYWORDS

気象時系列 (Meteorological Time Series) : 自己回帰モデル (Autoregressive Model AR Model) : 情報量規準 (Information Criterion IC) : AIC 最小化推定法 (Minimum AIC Estimation MAICE) : 予測誤差の期待値 (Final Prediction Error FPE) : マルコフ過程 (Markov Process) : 漸近的正規性 (Asymptotic Normality) : ケチの原理 (Principle of Parsimony)

## I 序 論

日本の地形は複雑で、このために気象要素の地域分布も多様である。しかも、特に山地部においては観測所が少なく実測資料も乏しい。そのため気象時系列の解析を試みる場合、まず既存のデータに基づき何等かの首肯できる根拠を持って、対象とする気象要素の時系列構造や空間的な相関構造を推定することが、その第一段階となる。この時系列構造や空間構造を明確にしておくことは、観測網の配置、未観測地点値の推定、欠測値の推定等の上で重要である<sup>1-2)</sup>。

統計モデルを導入して、気象時系列の構造規定、予測問題等を論ずる場合において、最も基本的で実用的なモデルが自己回帰モデル (Autoregressive Model : AR モデル) である<sup>3-4)</sup>。

AR モデルは時系列の現在の値を、過去の値に線形に依存する部分とそれでは表現できない部分の和として表現するモデルで、時系列が自己回帰過程に従う場合には、自己相関関数やスペクトル密度関数などの推定量として非常に良い性質を持つ最尤推定量が得られることが知られている<sup>5-6)</sup>。ところが、実際の問題においては AR モデルのパラメータ数は未知な場合が多く、複数のモデルが想定される。つまり、何次の AR モデルをあてはめて良いかわからないことが多い。この場合には、まず次数を推測してから分析する必要がある。

それでは、この次数をいかにして決定すれば良いであろうか？

これについては現在いくつかの方法が提案されているが、統一的なものは示されていない<sup>7-8)</sup>。

T. W. Anderson<sup>9)</sup> によるものは、可能性のある最低次数と最高次数を予め定めておき、その間のどの次数であるか定めるのを多重決定の問題として取り扱う方法であり、このために有意水準など予め多くのパラメータを定めておかなければならず、実用上の煩雑さを伴う。これとは異なる立場で、赤池<sup>10-12)</sup> は最尤法であてはめた回帰モデルによって予測した時の平均 2 乗予測誤差を見積る量として、予測誤差の期待値 FPE (Final Prediction Error) を示し、FPE の最小値を与えるモデルを最適なモデルとして採用する方法を提案してい

る。この統計量は、Kullback により定義された情報量とも関係を持つ<sup>13)</sup>。

ところで、著者等はこれまでの一連の研究で多項式モデルあるいはフーリエモデルによる地形データの構造決定における適合度検定の問題を論点として、従来の統計的仮説検定にかわる新たな統一的モデル評価規準として、赤池が提案している情報量規準 AIC が、この問題に対する極めて有効な接近法であることを数値実験、諸適用事例で論証してきた<sup>14)</sup>。

本稿のテーマである「気象要素の時系列構造と自己回帰モデルによる精度予測」の問題も、これらの延長線上にある。従って、気象時系列の構造をモデルによって把握することを最終的なねらいとしているが、そのための接近法は、モデルの構成と情報量による評価という一貫した視点に立っている<sup>15)</sup>。

## II 情報量規準 (Information Criterion: IC) とモデル評価

統計的なデータに関する命題あるいは仮説は、観測データの統計的分布に関する規定として表現される。ところが、観測値の“真の分布”は、観測値を人工的に生成する場合以外、正確に知られていることはない。データの統計的分布の取扱いに際しては、その分布をある程度近似するようなモデルを規定して進む以外に方法はない。そこで、種々の仮説設定は種々のモデルを想定することに対応する。そうして、最終的にはこれらの中から単一モデルの選択が要求される。ここに、統計モデル選択の理論的な研究の必要性が生じる。

これらの統計モデルの評価・選択という観点から、AIC をはじめとする各種の IC が提案されかなりの効果を上げている。ところが、この IC のふるまいについてある程度考察が行なわれているものもあるが、未知なものも多い。本来、各 IC はそれ固有の目的のために提案された統計量であるので、統計モデルの選択に際して一律に IC を導入することは充分留意しなければならない<sup>16-17)</sup>。

そこで、我々は AR モデルの評価・選択に(1)式で定式化された IC を導入して、各 IC を仮説検定と予測という両面から評価することにした。従って、一概に良悪は判定できないが、与えられた評価方式下ではどの IC がどの条件下で better であるかは議論することができる。

$$\begin{aligned} \text{AIC: } \text{AIC}(m) &= n \log R_{ssm} + 2m \\ \text{BIC: } \text{BIC}(m) &= n \log R_{ssm} + m \log n \\ \text{CAI: } \text{CAI}(m) &= n \log R_{ssm} + \frac{2n(m+1)}{n-m-2} \end{aligned} \quad (1)$$

ただし

$$\begin{aligned} R_{ssm} &: \text{Model}(m) \text{ の残差分散} & n &: \text{データ数} \\ R_{ssm} &= \frac{1}{n} \sum_{t=1}^n (x_t - \sum_{i=1}^m a_i x_{t-i})^2 & m &: \text{自由パラメータ数} \end{aligned}$$

### II-1 AR モデルにおける AIC<sup>18-20)</sup>

一変量の時系列  $\{x_t; t=1, \dots, n\}$  を考え、AR モデルを(2)式で定義する。

$$\text{Model(AR}(m)) : x_t = \sum_{i=1}^m a_i x_{t-i} + \varepsilon_t \quad (2)$$

$m$  は次数、 $a_i$  は自己回帰係数で、 $\varepsilon_t$  は平均 0、分散  $\sigma^2$  の正規分布とする。換言すれば、過去の値  $x_{t-m}, \dots, x_{t-1}$  が与えられた時には、 $x_t$  は平均  $a_1 x_{t-1} + \dots + a_m x_{t-m}$ 、分散  $\sigma^2$  の正規分布に従う。ここで、 $x_{1-m}, \dots, x_0$  は既知とし  $x_1, \dots, x_n$  に関する尤度を考えると、

$$L(a_1, \dots, a_m, \sigma^2) = f(x_1, \dots, x_n | x_{1-m}, \dots, x_0)$$

$$= \prod_{i=1}^n f(x_i | x_{i-m}, \dots, x_{i-1}) \quad (3)$$

となる。 $f(x_i | x_{i-m}, \dots, x_{i-1})$  は  $x_{i-m}, \dots, x_{i-1}$  が既知の下での  $x_i$  の条件付分布であり、平均  $a_1 x_{i-1} + \dots + a_m x_{i-m}$ 、分散  $\sigma^2$  の正規分布の密度関数であるから、

$$f(x_i | x_{i-m}, \dots, x_{i-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} \left( x_i - \sum_{i=1}^m a_i x_{i-i} \right)^2 \right\} \quad (4)$$

となり、**Model(AR(m))** の尤度は(5)式で表わされる。

$$L(a_1, \dots, a_m, \sigma^2) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( x_i - \sum_{i=1}^m a_i x_{i-i} \right)^2 \right\} \quad (5)$$

(5)式の対象数をとって、対象数尤度  $l$  は

$$l(a_1, \dots, a_m, \sigma^2) = -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n \left( x_i - \sum_{i=1}^m a_i x_{i-i} \right)^2 \quad (6)$$

となる。 $a_1, \dots, a_m$  と  $\sigma^2$  の最尤推定量を求めるには、この対象数尤度の一階微分を0とおいて、

$$\begin{aligned} \frac{\partial l}{\partial a_1} &= \frac{1}{\sigma^2} \sum_{i=1}^n x_{i-1} \left( x_i - \sum_{i=1}^m a_i x_{i-i} \right) = 0 \\ &\vdots \\ \frac{\partial l}{\partial a_m} &= \frac{1}{\sigma^2} \sum_{i=1}^n x_{i-m} \left( x_i - \sum_{i=1}^m a_i x_{i-i} \right) = 0 \\ \frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n \left( x_i - \sum_{i=1}^m a_i x_{i-i} \right)^2 = 0 \end{aligned} \quad (7)$$

を解けば良い。従って、(8)式の正規方程式解として、最尤推定量  $\hat{a}_1, \dots, \hat{a}_m$  が得られる。

$$\begin{bmatrix} C(1, 1) & \dots & C(1, m) \\ \vdots & & \vdots \\ C(m, 1) & \dots & C(m, m) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} C(1, 0) \\ \vdots \\ C(m, 0) \end{bmatrix} \quad (8)$$

$$\text{ただし, } C(i, j) = \sum_{i=1}^n x_{i-i} x_{i-j}$$

$\sigma^2$  の最尤推定量  $\hat{\sigma}^2$  は、

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n \left( x_i - \sum_{i=1}^m \hat{a}_i x_{i-i} \right)^2 \\ &= \frac{1}{n} \left( C(0, 0) - \sum_{i=1}^m \hat{a}_i C(i, 0) \right) \end{aligned} \quad (9)$$

となり、これを(6)式に代入して最大対象数尤度は

$$l(\hat{a}_1, \dots, \hat{a}_m, \hat{\sigma}^2) = -\frac{n}{2} \log 2\pi\hat{\sigma}^2 - \frac{n}{2} \quad (10)$$

となる。この  $m$  次の **AR** モデルは  $m+1$  個の自由パラメータを持っているので、その **AIC** は

$$\begin{aligned} \text{AIC}(m) &= -2l(\hat{a}_1, \dots, \hat{a}_m, \hat{\sigma}^2) + 2(m+1) \\ &= n(\log 2\pi + 1) + n \log \hat{\sigma}^2 + 2(m+1) \end{aligned} \quad (11)$$

により定義される。ただし、右辺第1項は次数に関係しない定数であるので省略でき、この場合(11)式は(1)式で定義した **AIC** になる。

この様にして推定された **MAICE** モデルは、ただちに予測に利用できる。現時点  $x_n$  までの時

系列が与えられた場合には、一時期先の予測値は

$$\hat{x}_{n+1} = \sum_{i=1}^m a_i x_{n+1-i} \quad (12)$$

となる。正確には  $x_{n+1}$  は平均  $\hat{x}_{n+1}$ 、分散  $\hat{\sigma}^2$  の正規分布に従うといえる。さらに、この  $\hat{x}_{n+1}$  から  $x_{n+2}$  の予測値は

$$\hat{x}_{n+2} = a_1 \hat{x}_{n+1} + \sum_{i=2}^m a_i x_{n+2-i} \quad (13)$$

となり、この予測値を代入して順次将来の予測値が求まる。

## II-2 正規分布モデルにおける AIC<sup>21)</sup>

正規分布モデルの同一性の判定をするために、条件付正規分布モデルの AIC を考える。2 組のデータ  $\{\theta_k(i) : i=1, \dots, nk\}$ ,  $\{\theta_l(i) : i=1, \dots, nl\}$  が違う正規分布に従うと仮定して(14)式 of モデルを考える。

$$\text{Model}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : f(\theta | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} \quad (14)$$

ここで  $(\mu, \sigma^2) = \begin{cases} \mu_1, \sigma_1^2 & \theta \text{ がデータ } \theta_k \text{ に属する場合} \\ \mu_2, \sigma_2^2 & \theta \text{ がデータ } \theta_l \text{ に属する場合} \end{cases}$

もし、データ  $\theta_k$  と  $\theta_l$  が同じ正規分布に従うなら  $\mu_1 = \mu_2 = \mu$ ,  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  が成立し、 $\theta$  がどちらに属していても、

$$\text{Model}(\mu, \mu, \sigma^2, \sigma^2) : f(\theta | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} \quad (15)$$

が成立する。このモデルは  $\text{Model}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$  に  $\mu_1 = \mu_2$ ,  $\sigma_1^2 = \sigma_2^2$  なる制約を課した自由パラメータ数 2 の条件付分布モデルと見做せる。 $\text{Model}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$  を仮定した時の対数尤度は(16)式となる。

$$\begin{aligned} l(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) &= \sum_{i=1}^{nk} \log f(\theta_k(i) | \mu_1, \sigma_1^2) + \sum_{i=1}^{nl} \log f(\theta_l(i) | \mu_2, \sigma_2^2) \\ &= -\frac{nk}{2} \log 2\pi - \frac{nk}{2} \log \sigma_1^2 - \frac{1}{2\sigma_1^2} \left( \sum_{i=1}^{nk} \theta_k(i)^2 - 2\mu_1 \sum_{i=1}^{nk} \theta_k(i) + nk\mu_1^2 \right) \\ &\quad - \frac{nl}{2} \log 2\pi - \frac{nl}{2} \log \sigma_2^2 - \frac{1}{2\sigma_2^2} \left( \sum_{i=1}^{nl} \theta_l(i)^2 - 2\mu_2 \sum_{i=1}^{nl} \theta_l(i) + nl\mu_2^2 \right) \\ &= -\frac{n}{2} \log 2\pi - \frac{nk}{2} \log \sigma_1^2 - \frac{1}{2\sigma_1^2} \left\{ \sum_{i=1}^{nk} \theta_k(i)^2 - 2\mu_1 \sum_{i=1}^{nk} \theta_k(i) + nk\mu_1^2 \right\} \\ &\quad - \frac{nl}{2} \log \sigma_2^2 - \frac{1}{2\sigma_2^2} \left\{ \sum_{i=1}^{nl} \theta_l(i)^2 - 2\mu_2 \sum_{i=1}^{nl} \theta_l(i) + nl\mu_2^2 \right\} \end{aligned} \quad (16)$$

ただし

$$n = nk + nl \quad \theta_k = \sum_{i=1}^{nk} \theta_k(i) \quad \theta_l = \sum_{i=1}^{nl} \theta_l(i)$$

対数尤度を  $(\mu_1, \sigma_1^2)$  に関して最大化するには、(16)式の最右辺 2 行目を考えれば良いから、最尤推定量は

$$\hat{\mu}_1 = \frac{\theta_k}{nk}, \quad \hat{\sigma}_1^2 = \frac{1}{nk} \sum_{i=1}^{nk} \theta_k(i)^2 - \hat{\mu}_1^2 \quad (17)$$

となる。同様に  $(\mu_2, \sigma_2^2)$  の最尤推定量は、最右辺 3 行目の最大化により

$$\hat{\mu}_2 = \frac{\theta_l}{nl}, \quad \hat{\sigma}_2^2 = \frac{1}{nl} \sum_{i=1}^{nl} \theta_l(i)^2 - \hat{\mu}_2^2 \quad (18)$$

となり、最大対数尤度は(19)式となる。

$$\begin{aligned} l(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2) &= -\frac{n}{2} \log 2\pi - \frac{nk}{2} \log \hat{\sigma}_1^2 - \frac{nl}{2} \log \hat{\sigma}_2^2 - \frac{nl}{2} \\ &= -\frac{n}{2} \log 2\pi - \frac{nk}{2} \log \hat{\sigma}_1^2 - \frac{nl}{2} \log \hat{\sigma}_2^2 - \frac{n}{2} \end{aligned} \quad (19)$$

**Model( $\mu, \mu, \sigma^2, \sigma^2$ )** を仮定すると、対数尤度が

$$\begin{aligned} l(\mu, \sigma^2) &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 \\ &\quad - \frac{1}{2\sigma^2} \left\{ \sum_{i=1}^{nk} \theta_k(i)^2 + \sum_{i=1}^{nl} \theta_l(i)^2 - 2\mu(\theta_k + \theta_l) + n\mu^2 \right\} \end{aligned} \quad (20)$$

となり、最尤推定量および最大対数尤度は(21), (22)式となる。

$$\hat{\mu} = \frac{\theta_k + \theta_l}{n}, \quad \hat{\sigma}^2 = \frac{1}{n} \left\{ \sum_{i=1}^{nk} \theta_k(i)^2 + \sum_{i=1}^{nl} \theta_l(i)^2 \right\} - \hat{\mu}^2 \quad (21)$$

$$l(\hat{\mu}, \hat{\sigma}^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \hat{\sigma}^2 - \frac{n}{2} \quad (22)$$

従って、**Model( $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ )**, **Model( $\mu, \mu, \sigma^2, \sigma^2$ )** の AIC は

$$\text{AIC}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = n \log 2\pi + nk \log \hat{\sigma}_1^2 + nl \log \hat{\sigma}_2^2 + n + 2 \times 4 \quad (23)$$

$$\text{AIC}(\mu, \mu, \sigma^2, \sigma^2) = n \log 2\pi + n \log \hat{\sigma}^2 + n + 2 \times 2 \quad (24)$$

となる。(23)式はデータ  $\theta_k(i)$  に **Model( $\mu_1, \sigma_1^2$ )** をあてはめた時の

$$\text{AIC}(\mu_1, \sigma_1^2) = nk \log 2\pi + nk \log \hat{\sigma}_1^2 + nk + 2 \times 2$$

とデータ  $\theta_l(i)$  に **Model( $\mu_2, \sigma_2^2$ )** をあてはめた時の

$$\text{AIC}(\mu_2, \sigma_2^2) = nl \log 2\pi + nl \log \hat{\sigma}_2^2 + nl + 2 \times 2$$

の和に等しいことがわかる。すなわち、

$$\text{AIC}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \text{AIC}(\mu_1, \sigma_1^2) + \text{AIC}(\mu_2, \sigma_2^2) \quad (25)$$

が成立する。

### II-3 FPE (Final Prediction Error) と $\text{AIC}^{(22-25)}$

最尤法であてはめた回帰モデルによって予測した時の、平均 2 乗予測誤差を見積る統計量として赤池の FPE がある。

$$\text{FPE} = \frac{(n+m)}{(n-m)} \hat{\sigma}^2 \quad (26)$$

$$\text{ただし } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left( x_i - \sum_{i=1}^m \hat{a}_i x_{i-i} \right)^2$$

これは、モデルをあてはめた時の説明変数の分布と、予測する時の説明変数の分布が一致している時の平均 2 乗予測誤差の推定量である。ここで、(26)式の対数をとってデータ数  $n$  をかけると

$$\begin{aligned} n \log \text{FPE} &= n \log \hat{\sigma}^2 + n \log \frac{n+m}{n-m} \\ &= n \log \hat{\sigma}^2 + n \log \left( 1 + \frac{2m}{n-m} \right) \\ &\approx n \log \hat{\sigma}^2 + 2m \end{aligned} \quad (27)$$

となるから、 $n$  が十分に大きい時(11)式の AIC と  $n \log \text{FPE}$  は定数を除いて近似的に等しいこと

がわかる。従って、モデルをあてはめる際に AIC でモデルを評価することは、平均 2 乗予測誤差を小さくするモデルを探すことに等しい。<sup>26-27)</sup>

### Ⅲ 気象時系列と基礎資料<sup>28-29)</sup>

気象要素の長期変動の問題、すなわち降水量や気温について周期性や経年的傾向があるのではないかという問題は、かなり以前から太陽の黒点周期（約11年）との関係等から論じられている。もちろん、地質学的に見れば永期、間永期の存在が示す長周期的変化があることは確かであるが、ここで問題としているのは5年とか10年が単位の気象変動である。

対象とする気象要素は、月雨量、月平均気温、日平均気温、降雪季の最大積雪深で、資料は京都大学農学部演習林気象報告芦生演習林観測所からとった。表—1 にその概要を示す。

表—1 気象要素と基礎資料概要  
Table 1. Outlines of applied meteorological elements and data sampling for modelling

Meteorological elements	Sampling term	Data length
Daily mean of air temperature	1983. Jan. 1 — 31 1983. Aug. 1 — 31 1984. Jan. 1 — 31	31
Annual record of maximum snow depth	1925. — 1984.	60
Monthly total precipitation	1969. Jan. — 1984. Jan.	181
Monthly mean of air temperature	1966. Jan. — 1984. Jan.	217

-Remarks-

Location of the observatory of Kyoto University Forest in Asyu is just where latitude 35.18 N meets longitude 135.43E, with 359 meters above the sea level.

The average annual temperature is 13.5°C with the maximum 39.0°C and minimum -19.5°C; the precipitation is more than 2500mm, and especially there is heavy snowfall amounting 1 to 1.5 meters on the level in winter.

### Ⅳ 解析結果と考察

#### Ⅳ—1 日平均気温の時系列構造と予測モデル

1983年8月1日～31日の日平均気温のデータを用いて、この月のある日までの記録に基づいて、翌日の平均気温を予測する AR モデルを考える。ただし、AR モデルの最高次数は10次までとする。表—2 より、4 次の AR モデルが MAICE モデルとなり、いずれの IC もこのモデルを最良なモデルとして選択している。このモデルを用いて、「5日目」以降の平均気温を予測してプロットしたのが図—1 である。「9日目」までのあてはまりは良好であるとはいえないが、それ以降ではかなり再現されている。



表-2 MAICE による1983年8月の日平均気温の予測モデルの決定

Order: 次数 IC: AIC, BIC, CAI Variance: 残差分散

Table 2. Order determination and estimation of AR models performed by the MAICE procedure for prediction of daily mean of air temperature in Aug. 1983

ORDER	AIC	BIC	CAI	VARIANCE
0	97.32	98.76	99.75	0.126759D+01
1	87.02	89.89	89.91	0.852333D+00
2	86.38	90.68	89.92	0.782654D+00
3	85.15	90.89	89.55	0.705343D+00
4	<u>81.65</u>	<u>88.82</u>	<u>87.15</u>	<u>0.590685D+00</u>
5	82.61	91.22	89.48	0.571174D+00
6	84.12	94.16	92.67	0.562249D+00
7	86.08	97.55	96.65	0.561418D+00
8	84.80	97.70	97.80	0.505023D+00
9	86.79	101.13	102.68	0.504921D+00
10	84.55	100.33	103.89	0.440398D+00

MAICE Order=4 Minimum AIC=81.65

Residual Variance=0.59068D+00

AR Model:

$$X_n = 0.5984X_{n-1} - 0.6989X_{n-2} + 0.5763X_{n-3} - 0.4015 + \varepsilon_n$$

$$\varepsilon_n \approx N(0, 0.590)$$

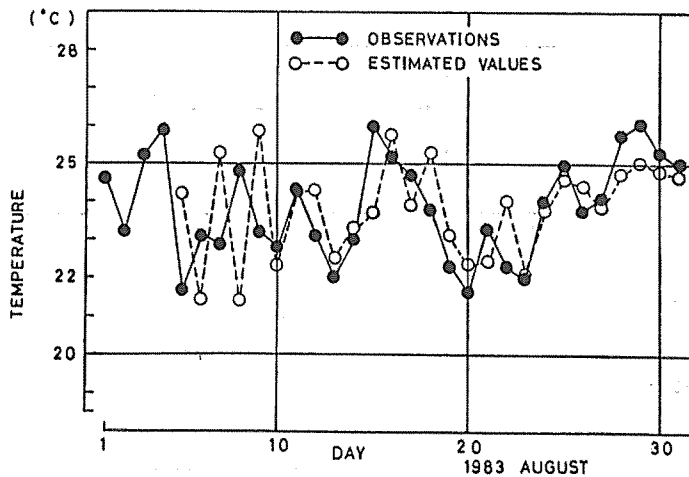


図-1 Model (4) による1983年8月の各日の平均気温の予測 ●: 観測値 ○: 予測値  
Figure 1. Plot of predicted values for daily mean of air temperature in Aug. 1983 by Model (4)

同様に、1983年1月1日～31日のデータに対して AR モデルをあてはめ、このモデルを用いて翌年1984年1月の日平均気温を予測する。表-3より、1次のモデルのあてはまりが最も良いことがわかり、各 IC は同一モデルを選択している。図-2は、このモデルから「2日目」以降の気温を予測したもので、全体的な傾向は比較的良く近似されている。この場合の自己相関関数は単純マルコフ型となる。

これらの結果は、夏季と冬季における日平均気温の時系列構造に差異があり、夏季の気温にはある期間の経日的持続性があるのに対し、冬季では前日のみの気温変化に支配される傾向があることを示している。また、AR モデル選択上の各 IC の評価は、まったく一致する。

表-3 MAICE による1983年1月の日平均気温の予測モデルの決定  
 Table 3. Order determination and estimation of AR models performed by the MAICE procedure for prediction of daily mean of air temperature in Jan. 1983

ORDER	AIC	BIC	CAI	VARIANCE
0	119.04	120.47	121.47	0.255388D+01
<u>1</u>	<u>112.08</u>	<u>114.95</u>	<u>114.97</u>	<u>0.191303D+01</u>
2	113.56	117.86	117.10	0.188091D+01
3	115.56	121.29	119.96	0.188082D+01
4	116.83	124.00	122.33	0.183693D+01
5	117.44	126.04	124.31	0.175642D+01
6	119.42	129.46	127.97	0.175580D+01
7	121.29	132.76	131.86	0.174806D+01
8	122.26	135.17	135.26	0.169124D+01
9	124.26	138.60	140.16	0.169106D+01
10	124.02	139.80	143.36	0.157336D+01

MAICE Order=1 Minimum AIC=112.08

Residual Variance=0.191303D+01

AR Model:

$$X_n = 0.5008X_{n-1} + \varepsilon_n$$

$$\varepsilon_n \sim N(0, 1.913)$$

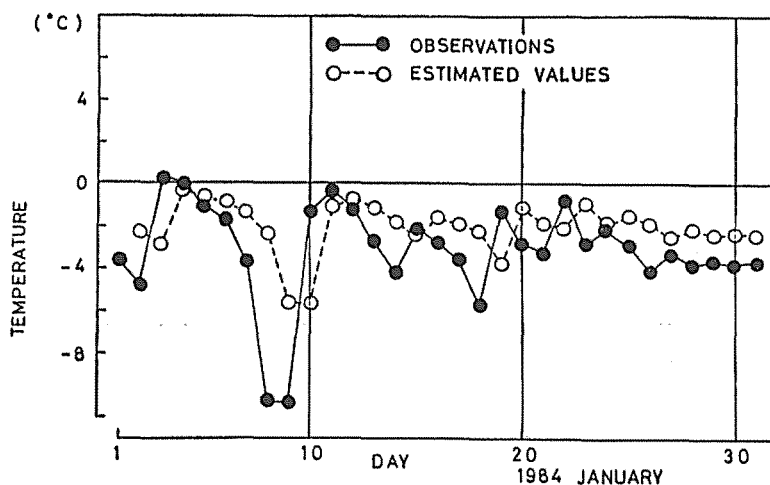


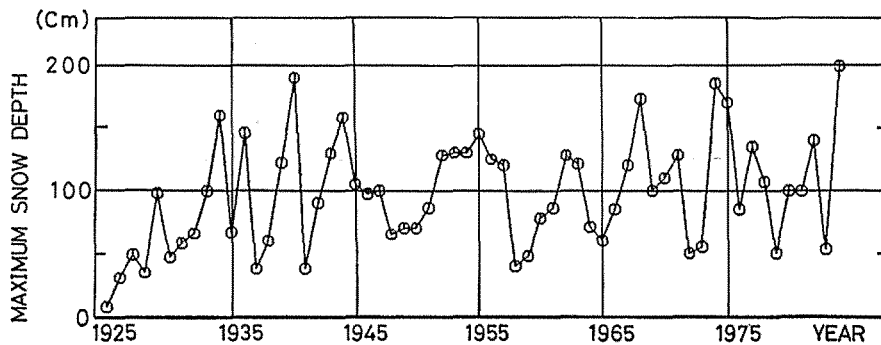
図-2 Model (1) による翌年 (1984年) 1月の各日の平均気温の予測  
 Figure 2. Plot of predicted values for daily mean of air temperature in Jan. 1984 by Model (1)

#### IV-2 降雪季の最大積雪深の予測モデル<sup>30)</sup>

1925年～1984年の降雪季の最大積雪深のデータに対する AR モデルを考える。図-3は最大積雪深をプロットしたもので、降雪季の最大積雪月の出現頻度が2月に集中していることがわかる。表-4に、AR モデルをあてはめた結果を示す。AIC と CAI では5次の AR モデル、BIC では0次のモデルが最良と判断されている。0次のモデルとは、時系列構造にいかなる周期性・経年的傾向も認められず、母集団特性がアンサンブル平均と分散で示される、一種の正規分布モデルを意味する。

Month	Frequency	Relative Freq. (Percent)
11	0	0
12	3	4.76
1	15	23.81
2	38	60.32
3	7	11.11
4	0	0

1 Season : November - April



図一3 1925～1984年の降雪季における最大積雪深月の頻度と最大積雪深  
Figure 3. Plot of annual record of maximum snow depth  
in winter from 1925 to 1984

表一4 MAICE による降雪季の最大積雪深の予測モデルの決定

Table 4. Order determination and estimation of AR models performed by the MAICE procedure for prediction of annual record of maximum snow depth in winter season

ORDER	AIC	BIC	CAI	VARIANCE
0	602.96	<u>605.05</u>	605.17	0.131045D+04
1	604.28	608.47	606.71	0.129563D+04
2	604.08	610.36	606.81	0.124902D+04
3	605.97	614.34	609.08	0.124669D+04
4	607.95	618.42	611.53	0.124629D+04
5	<u>600.38</u>	612.95	<u>604.53</u>	<u>0.106258D+04</u>
6	600.48	615.14	605.30	0.102945D+04
7	602.44	619.20	608.04	0.102884D+04
8	604.13	622.97	610.61	0.102339D+04
9	605.91	626.86	613.41	0.101981D+04
10	607.87	630.91	616.51	0.101906D+04
11	608.92	634.05	618.83	0.100299D+04
12	610.65	637.87	621.98	0.998457D+03
13	611.61	640.93	624.52	0.981332D+03
14	613.57	644.98	628.22	0.980672D+03
15	614.40	647.90	630.97	0.961709D+03

MAICE Order=5 Minimum AIC=600.38

Residual Variance=0.106258D+04

AR Model:

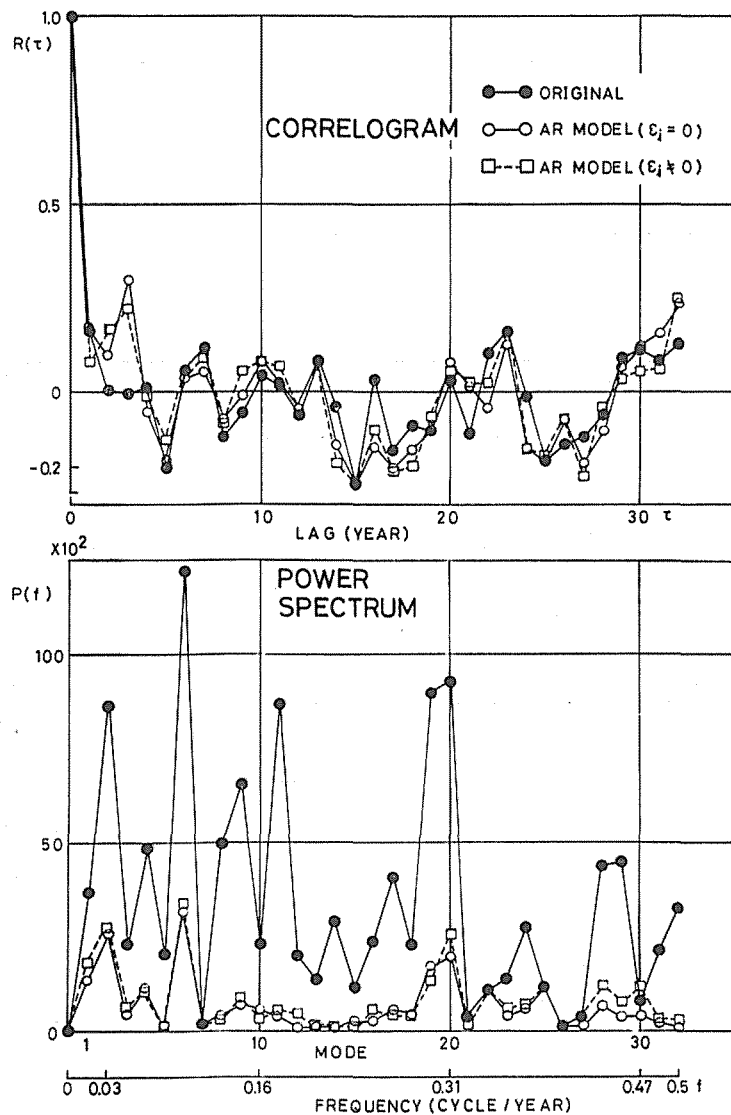
$$X_n = 0.1034X_{n-1} - 0.1769X_{n-2} - 0.1779X_{n-3} + 0.0889X_{n-4} - 0.4108X_{n-5} + \varepsilon_n$$

$$\varepsilon_n \sim N(0, 1062.58)$$

図一4に原データ、5次のARモデル( $\varepsilon_i=0$ )、5次のARモデル+乱数( $\varepsilon_i \neq 0$ )の各系列についてのコレログラム・パワースペクトルを示す。

ただし、原データとARモデルのデータ数が異なり、計算処理もFFTアルゴリズムを使用しているので当然グラフの縦軸のスケールに若干の差が生じる。

これを見ると、顕著ではないがlagにして7年程度の周期的傾向が認められ、またARモデルおよびこれに乱数を加えてシミュレーションした擬似系列は、原データの特徴にかなり近づけられたものであることがわかる。これらの結果から、最大積雪深の予測モデルとしては、AIC, CAIの選択した5次のARモデルを採用することが肯定される。



図一4 Model (5) :  $\varepsilon_i=0$ , Model (5)+乱数 :  $\varepsilon_i \neq 0$ , 観測データからのコレログラムとパワースペクトルの比較

● : 観測データ ○ : Model (5) □ : Model (5)+乱数

Figure 4. Comparison of spectral qualities of the following time series data; original data, Model (5) and simulated Model (5)

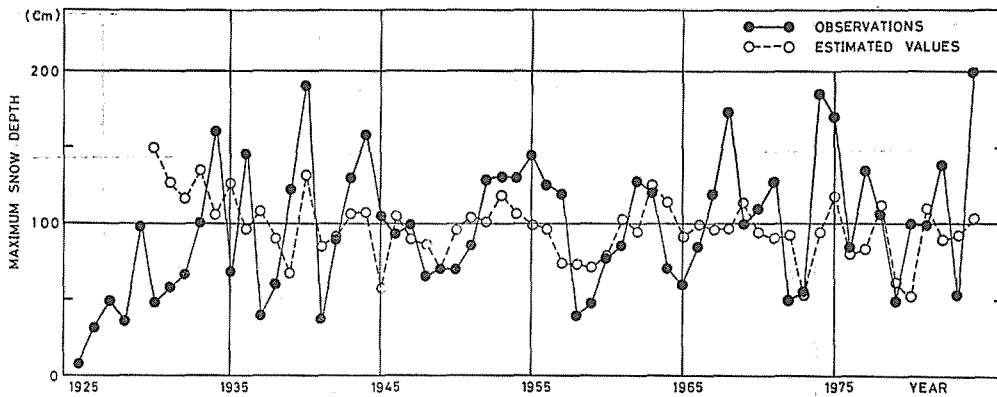


図-5 Model(5) による各年の最大積雪深の予測 ●: 観測値 ○: 予測値  
Figure 5. Plot of predicted values for annual record of maximum snow depth by Model(5)

予測モデル

$$x_n = 0.1034x_{n-1} - 0.1769x_{n-2} - 0.1779x_{n-3} + 0.0689x_{n-4} - 0.4108x_{n-5} + \varepsilon_n$$

$$\varepsilon_n \approx N(0, 1062.58)$$

このモデルを用いて、「1930年」以降の最大積雪深を予測してプロットしたものが図-5で、分布変動が大きい部分を除けば、全体的には比較的再現されていると思われる。

#### IV-3 月雨量の時系列構造と予測モデル<sup>31-35)</sup>

雨量系列の周期性・経年的傾向は、梅雨季および台風季の2回の雨季、さらに冬の降雪季を前提として考えることができる。

菅原・赤池の「水資源の変動様相に関する調査報告」<sup>36)</sup>、菅原・勝山の「気象資料の長期変動」<sup>37)</sup>の研究によると、…脊梁山脈にさえぎられて台風の影響を受けにくく、冬期に季節風により地形的降雪のある裏日本の山腹では、雨量は極めて規則的な季節変化を示すこともあるが、一般的には、雨量系列には季節変化以外のいか

## METEOROLOGICAL ELEMENT:

### MONTHLY TOTAL PRECIPITATION

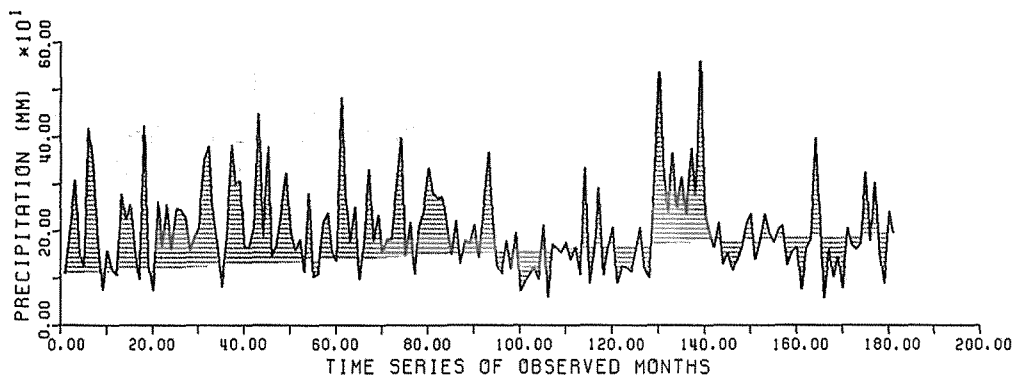
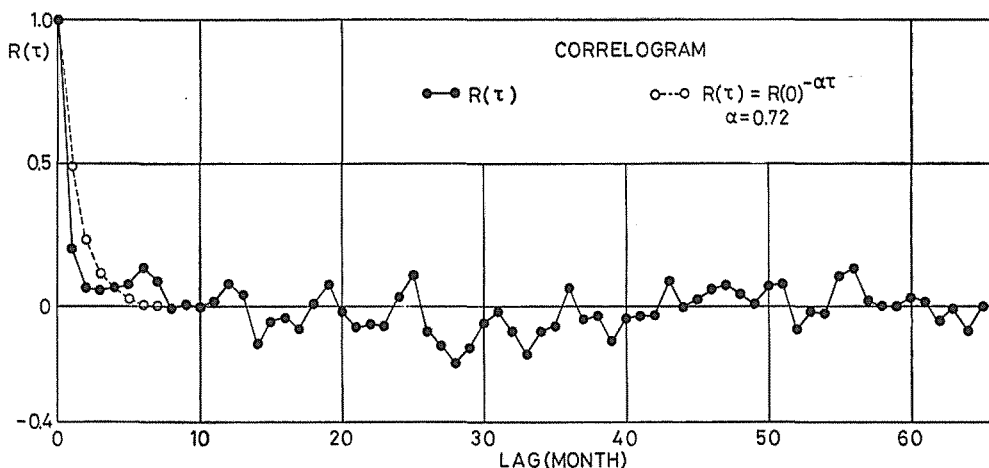


図-6 1969年1月～1984年1月の月別降水量  
Figure 6. Plot of monthly total precipitation from Jan. 1969 to Jan. 1984



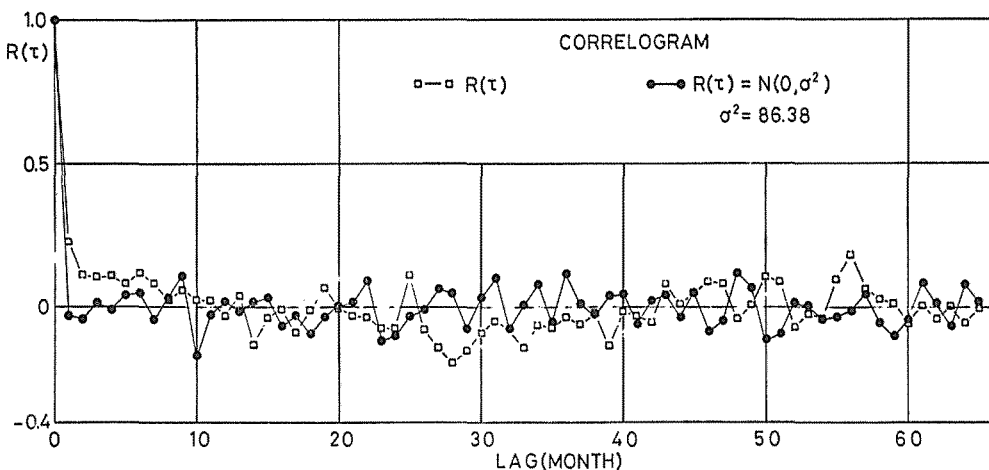
図一7 月別降水量から算定した自己相関関数のコレログラムと相関の通減率を  $\alpha=0.72$  とした理論的な単純マルコフ型に従うコレログラムの比較  
 ●:  $R(\tau)$  ○:  $R(\tau) = R(0)^{-\alpha\tau}$   $\alpha=0.72$

Figure. 7. Comparison of correlograms between monthly total precipitation and Markov process derived from theoretical formula as the diminishing ratio is 0.72

なる周期性も傾向も認められない… ことが報告されている。しかし、この結果に対していくつかの異議も出されている。

さて、図一6は1969年1月～1984年1月の月別降水量、図一7\*はこれより算定した自己相関関数のコレログラムを示している。コレログラムは、半年および年周期を示しているが全体的には極めて雑音的であり、全分散の約85%が雑音部分である。ちなみに、この値は先の菅原・勝山の報告にある「鳥取県境港」のケースとほぼ同値を示しており、京都では約60%の値が示されている。

そこで、この雑音部分の性質を調べるために、月資料から1月、2月、…、12月のそれぞれの



図一8 季節変化のパターンを除去した雨量系列のコレログラムと乱数から生成した擬似系列のコレログラムの比較 □: 雨量系列 ●: 擬似系列

Figure 8. Comparison of correlograms between time series of monthly total precipitation eliminated the effect of seasonal change and the simulated Gaussian model

月別の累年平均値を引き去った時系列を作る。つまり、季節変化のパターンからのずれの部分だけを取り出す。次に、この時系列と等しい分散を持つ擬似系列を乱数から生成し、そのコレログラムを比較した結果が図-8である。両者の差はほとんど認められない。すなわち、乱数でもこの程度の変動が出るのであるから、月別雨量から季節変動の型を取り去った残りはいかなる周期性も持たない不規則変動であると考えられ、この結果は菅原等の報告と一致する。

この雨量時系列に対して、MAICE モデルを判定したのが表-5である。AIC, CAI はいずれも2次, BIC は1次のAR モデルを選択しており、コレログラムから検討した結果を裏付けて

表-5 MAICE による月別降水量の予測モデルの決定  
Table 5. Order determination and estimation of AR models performed by the MAICE procedure for prediction of monthly total precipitation

ORDER	AIC	BIC	CAI	VARIANCE
0	2079.30	2082.50	2081.36	0.564713D+04
1	2066.67	2073.07	2068.81	0.520878D+04
2	2066.15	2075.74	2068.38	0.513667D+04
3	2066.85	2079.64	2069.19	0.509983D+04
4	2068.61	2084.60	2071.09	0.509311D+04
5	2070.57	2089.76	2073.22	0.509215D+04
6	2070.72	2093.11	2073.56	0.504041D+04
7	2072.68	2098.27	2075.74	0.503928D+04
8	2073.23	2102.02	2076.52	0.499892D+04
9	2074.64	2106.62	2078.20	0.498266D+04
10	2076.58	2111.76	2080.44	0.498103D+04
11	2076.99	2115.38	2081.17	0.493760D+04
12	2077.64	2119.22	2082.17	0.490070D+04
13	2078.82	2123.60	2083.73	0.487859D+04
14	2072.59	2120.57	2077.91	0.466187D+04
15	2074.51	2125.69	2080.27	0.465982D+04
16	2076.32	2130.70	2082.55	0.465494D+04
17	2078.32	2135.90	2085.04	0.465488D+04
18	2079.72	2140.49	2086.97	0.463936D+04
19	2080.48	2144.45	2088.29	0.460767D+04
20	2082.45	2149.62	2090.85	0.460692D+04
21	2084.19	2154.56	2093.22	0.460039D+04
22	2085.01	2158.58	2094.71	0.457058D+04
23	2086.97	2163.73	2097.35	0.456935D+04
24	2087.85	2167.81	2098.97	0.454125D+04
25	2089.62	2172.78	2101.50	0.453551D+04
26	2085.71	2172.07	2098.40	0.438990D+04
27	2084.35	2173.91	2097.88	0.430912D+04
28	2081.60	2174.36	2096.00	0.419747D+04
29	2082.99	2178.94	2098.30	0.418326D+04
30	2084.99	2184.14	2101.26	0.418325D+04
31	2086.73	2189.09	2104.00	0.417743D+04
32	2088.58	2194.13	2106.88	0.417379D+04
33	2089.45	2198.19	2108.82	0.414779D+04
34	2090.43	2202.38	2110.93	0.412465D+04
35	2092.40	2207.54	2114.06	0.412385D+04
36	2092.66	2211.01	2115.54	0.408452D+04
37	2091.21	2212.76	2115.34	0.400740D+04
38	2093.12	2217.86	2118.55	0.400536D+04
39	2090.10	2218.04	2116.88	0.389585D+04
40	2091.81	2222.95	2119.99	0.388961D+04

MAICE Order=2 Minimum AIC=2066.15 Residual Variance=0.513667D+04  
AR Model:

$$X_n = 0.24496X_{n-1} + 0.1744X_{n-2} + \varepsilon_n$$

$$\varepsilon_n \approx N(0, 5136.67)$$

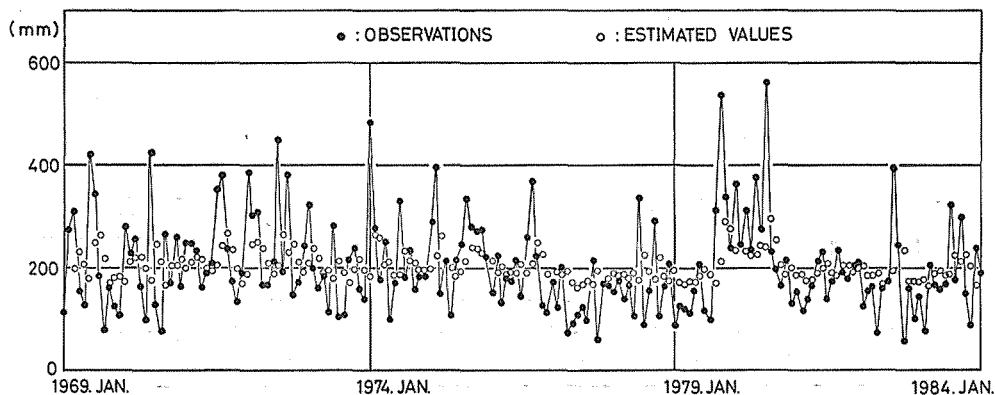


図-9 Model(2) による月別降水量の予測 ●: 観測値 ○: 予測値  
Figure 9. Plot of predicted values for monthly total precipitation by Model(2)

いる。IC 間の評価の差は、この場合ほとんど有意ではないと見做して良い。従って、雨量系列の自己相関関数は単純マルコフ型に近似されることが推測される。

\* 相関の遅減率を  $\alpha$ :  $\alpha = (1 - \rho) / \Delta t$ , 自己相関関数を  $R(\tau)$ :  $R(\tau) = R(0) e^{-\alpha \tau}$  で定義した単純マルコフ型のコレログラムを図-7中に示す

予測モデル

$$x_n = 0.24495x_{n-1} + 0.11744x_{n-2} + \varepsilon_n$$

$$\varepsilon_n \sim N(0, 5136.67)$$

図-9 は、このモデルを用いて月雨量を予測してプロットしたものである。

#### IV-4 月平均気温の時系列構造と予測モデル<sup>38)</sup>

図-10は1966年1月～1984年1月の月別平均気温、図-11はこの時系列のコレログラムを示し

## METEOROLOGICAL ELEMENT:

### MONTHLY MEAN OF AIR TEMPERATURE

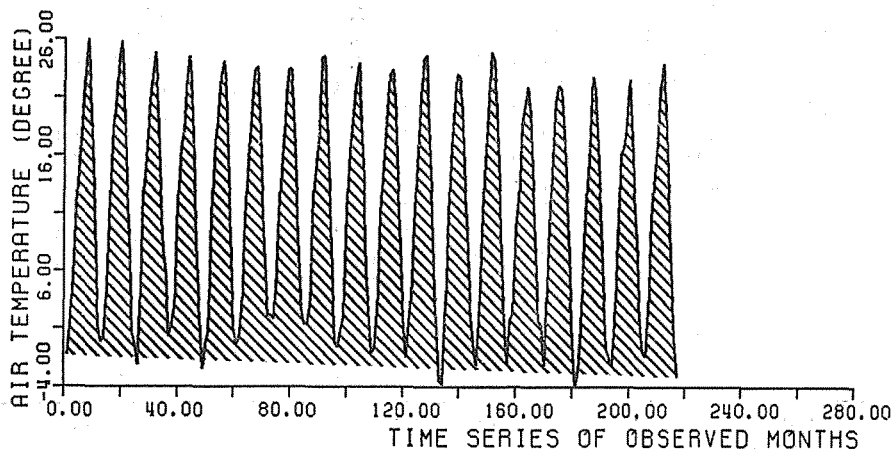


図-10 1966年1月～1984年1月の月別平均気温  
Figure 10. Plot of monthly mean of air temperature from Jan. 1966 to Jan. 1984



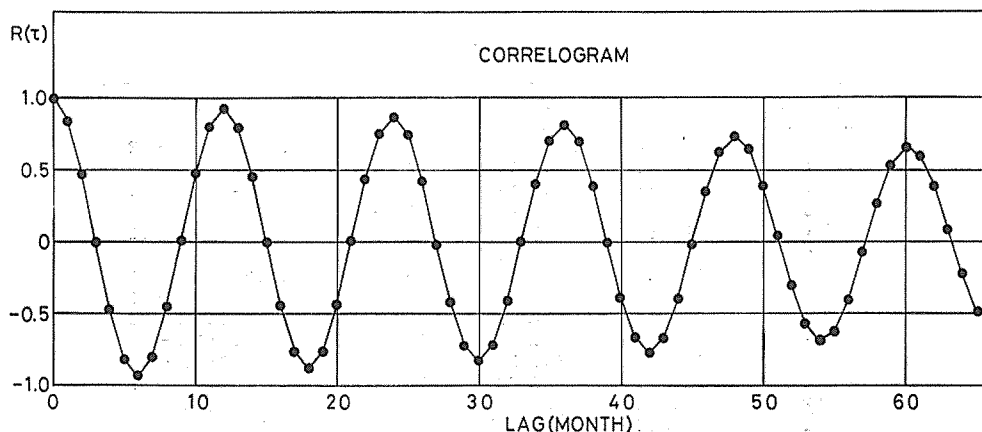


図-11 月別平均気温から算定したコレログラム  
Figure 11. Correlogram of monthly mean of air temperature

ており、規則的な年周期が現われている。これは、月雨量の場合とは著しく対称的で全分散のうち雑音部分の占める割合は、ほぼ10%程度である。

気温から季節変化（累年平均値）を除いた時系列と、乱数による擬似系列のコレログラムを比較したのが図-12で、周期的変化は認められないがある程度の持続性が現われている。**MAICE**モデルを判定した結果が表-6である。**IC** は、いずれも異なったモデルを選択している。

**AIC=40      BIC=12      CAI=35**

残差分散は、パラメータの増加に対して指数関数的に減少し、各 **IC** はこの残差分散を著しく減少させるモデルをいずれもの確に採用している。その意味においては、いずれの **IC** も反応性は良好である。**ケチの原理 (Principle of Parsimony)** に従うならば **BIC** の選択を採用することが考えられ、この場合、持続性は時定数で12ヶ月となる。しかし、図-12の時系列のコレログ

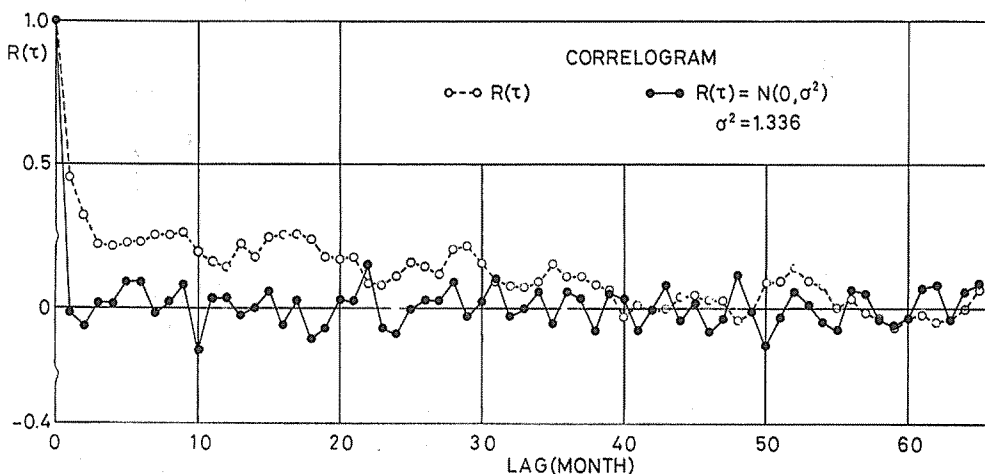


図-12 季節変化のパターンを除去した気温系列のコレログラムと乱数から生成した擬似系列のコレログラムの比較  
○: 気温系列    ●: 擬似系列  
Figure 12. Comparison of correlograms between time series of monthly mean of air temperature eliminated the effect of seasonal change and the simulated Gaussian model

表一6 MAICE による月別平均気温の予測モデルの決定  
 Table 6. Order determination and estimation of AR models performed by the MAICE procedure for prediction of monthly mean of air temperature

ORDER	AIC	BIC	CAI	VARIANCE
0	1474.34	1477.72	1476.39	0.517840D+02
1	1198.18	1204.94	1200.29	0.143715D+02
2	917.03	927.17	919.22	0.389782D+01
3	849.63	863.15	851.91	0.283087D+01
4	790.60	807.50	793.00	0.213689D+01
5	763.75	784.03	766.28	0.187086D+01
6	763.81	787.46	766.50	0.185419D+01
7	765.65	792.69	768.52	0.185290D+01
8	761.79	792.21	764.86	0.180353D+01
9	751.92	785.72	755.21	0.170751D+01
10	751.39	788.57	754.92	0.168768D+01
11	725.97	766.52	729.76	0.148734D+01
12	708.48	752.42	712.56	0.135961D+01
13	710.06	757.38	714.45	0.135697D+01
14	710.98	761.67	715.70	0.135021D+01
15	709.25	763.33	714.32	0.132722D+01
16	710.28	767.73	715.73	0.132128D+01
17	710.17	771.00	716.02	0.130849D+01
18	711.98	776.19	718.26	0.130735D+01
19	712.33	779.93	719.07	0.129748D+01
20	713.88	784.86	721.10	0.129480D+01
21	715.36	789.72	723.08	0.129171D+01
22	716.70	794.44	724.95	0.128779D+01
23	707.47	788.59	716.27	0.122280D+01
24	694.66	779.16	704.05	0.114214D+01
25	690.52	778.39	700.52	0.111025D+01
26	692.51	783.77	703.15	0.111025D+01
27	693.45	788.09	704.76	0.110484D+01
28	694.47	792.49	706.47	0.109984D+01
29	696.16	797.56	708.89	0.109828D+01
30	695.23	800.01	708.71	0.108355D+01
31	696.75	804.90	711.01	0.108114D+01
32	698.56	810.10	713.64	0.108022D+01
33	700.45	815.37	716.38	0.107968D+01
34	694.35	812.64	711.15	0.104008D+01
35	679.92	801.60	697.63	0.964249D+00
36	679.83	804.89	698.48	0.955020D+00
37	679.82	808.26	699.45	0.946219D+00
38	681.69	813.51	702.33	0.945658D+00
39	683.69	818.89	705.37	0.945658D+00
40	676.74	815.31	699.50	0.907418D+00
41	677.99	819.95	701.86	0.904304D+00
42	679.69	825.03	704.72	0.903058D+00
43	681.16	829.88	707.37	0.900860D+00
44	682.43	834.53	709.87	0.897831D+00
45	684.29	839.76	712.99	0.897232D+00
46	685.88	844.74	715.88	0.895559D+00
47	684.83	847.06	716.17	0.883031D+00
48	686.67	852.29	719.39	0.882399D+00
49	687.78	856.77	721.92	0.878767D+00
50	688.01	860.39	723.62	0.871660D+00
51	690.01	865.76	727.12	0.871633D+00
52	690.85	869.98	729.52	0.867000D+00
53	692.44	874.95	732.70	0.865345D+00
54	693.74	879.64	735.64	0.862589D+00
55	695.74	885.01	739.33	0.862577D+00
56	696.79	889.45	742.11	0.858825D+00
57	697.93	893.96	745.02	0.855406D+00
58	699.93	899.34	748.85	0.855399D+00
59	701.14	903.94	751.94	0.830599D+00
60	697.54	903.72	750.27	0.830599D+00

MAICE Order=40 Minimum AIC=676.74 Residual Variance=0.90742D+00

AR Model:

$$X_n = 0.41595X_{n-1} + 0.04759X_{n-2} + \dots + 0.06476X_{n-39} - 0.20648X_{n-40} + \varepsilon_n$$

$$\varepsilon_n \sim N(0, 0.90742)$$

ラムは、これよりも幾分長い持続性を持っているようであり、コログラムの0相関あるいはそれに近似する lag を基準とすれば、CAI, AIC の選択したモデルが採用されることになる。特に AIC の選択（時定数40ヶ月）は、厳密な意味での0相関に致る持続性という観点に立てば、最も妥当な結果となる。

予測モデル

$$x_n = 0.41595x_{n-1} + 0.04759x_{n-2} - \dots + 0.06476x_{n-39} - 0.20648x_{n-40} + \varepsilon_n$$

$$\varepsilon_n \sim N(0, 0.90742)$$

#### IV-5 IC の評価と正規性の同一性の判定

##### ① IC の評価と特性<sup>39-42)</sup>

(1)式で定義した IC 統計量は、右辺2項目の違いとして比較できる。BIC, CAI は2項目に自由パラメータ数  $m$  とデータ数  $n$  の2因子を含み、AIC は自由パラメータ数のみに依存している。そこで、データ数  $n$  を変化させた場合の2項目の挙動を次数との関係で示したのが図-13である。

BIC は、データ数が増加するにつれて2項目の傾きを増大させ AIC, CAI との差が著しくなる。低次でモデルが選択される場合は AIC, CAI に近い結果を得ることが推定されるが、高次においては2項目の影響に過大に反応し高次のモデル選択が困難となることがわかる。CAI はデータ数が増加すればするほど AIC に漸近し、同一モデルあるいは近似モデルを選択する傾向になる。ただし、CAI の2項目は双曲線であり  $-2n$ ,  $\frac{-2n(n-1)}{m-n+2}$  の漸近線を有しており、こ

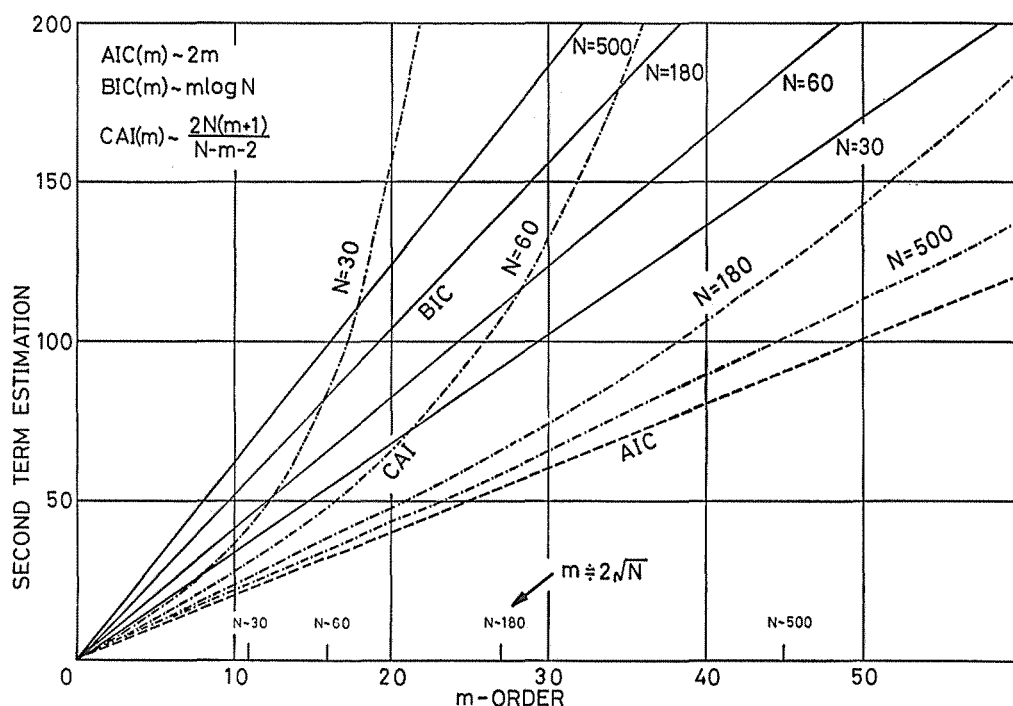


図-13 データ数の変化に対する IC の2項目の挙動と次数との関係 ( $N=30, 60, 180, 500$ )

$$AIC(m) \sim 2m \quad BIC(m) \sim m \log N \quad CAI(m) \sim \frac{2N(m+1)}{N-m-2}$$

Figure 13. Relationship between the second term estimation of the IC and order with the change of data length

のためにデータ数  $n$  に較べて自由パラメータが大きすぎるモデルを選択する危険性を回避している。赤池が AIC 適用上の範囲として示している  $2\sqrt{n}$  (高々  $n/2$ ) の基準は、パラメータの推定誤差に対する量尤推定量の漸近正規性に基づいたものであり、CAI は近似的にこれを満足していることが理解される。図-13 で示した  $n$  の変化に対する AIC の適用範囲は、CAI の安定した選択領域と極めて良く一致することがわかる。

これらのことより、データ数が比較的少ない場合には各 IC は同一モデルを選択することが期待される。データ数が増加すると各 IC は異ったモデルを選択する傾向が生じ、BIC は残差分散の減少傾向に対応して常に低次のモデルを選択し、2 項目のために高次モデルの分解能が悪くなる。CAI, AIC は高次でのモデルに対しても敏感に反応し、その適用領域が AIC では  $\sqrt{2n}$ , CAI では 2 項目の漸近線  $\frac{-2n(n-1)}{m-n+2}$  として control され、従って近似的なモデルを選択する。高次モデルの分解能という観点に立てば、

$$\text{BIC} \ll \text{CAI} < \text{AIC}$$

の関係が成立する。この関係を簡単な数値実験で示す。

正弦波+雑音の擬似系列  $x(t)$  を仮定し、正規乱数より生成したノイズレベルを  $\sigma_1^2=0.05 : \sigma_2^2=0.10 : \sigma_3^2=0.50$  の3段階として、500 個のデータを作り各ケースごとに AR モデルのあてはめを行う。擬似系列  $x(t)$  を図-14 に示す。

$$x(t) = \sin\left(\frac{\pi t}{15}\right) + \gamma(t)$$

$$\gamma(t) \approx 500(0, \sigma_n^2)$$

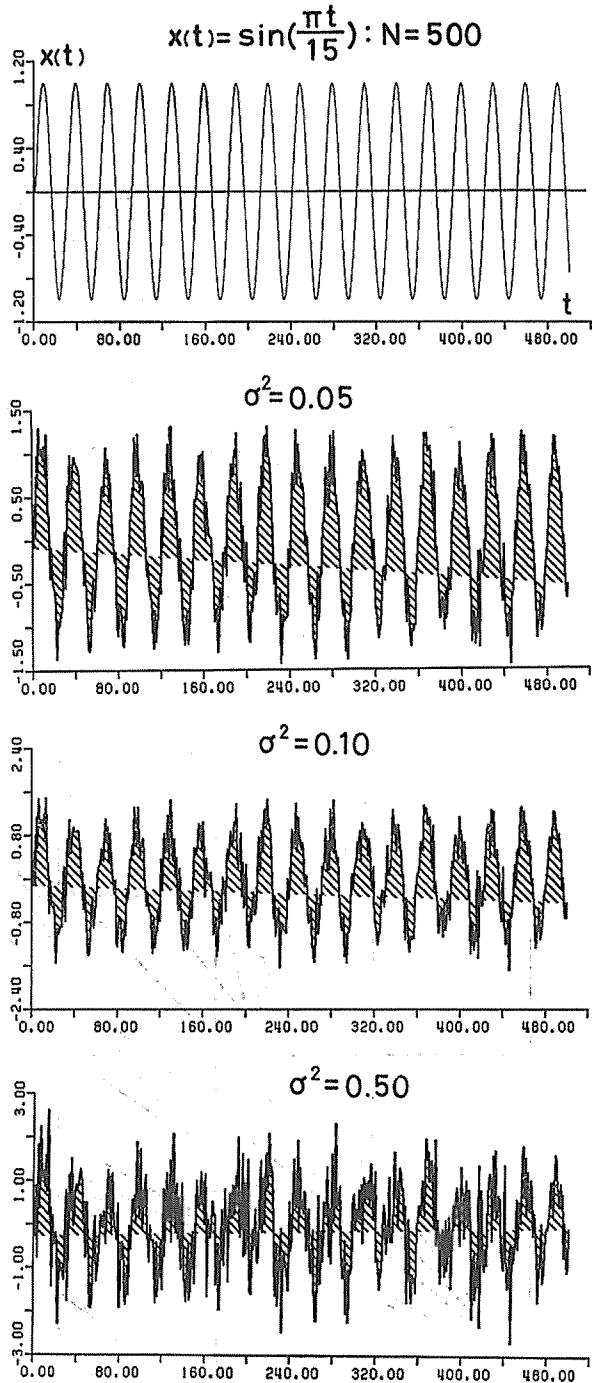


図-14 数値実験に用いた正弦波とノイズレベルを変化させた擬似系列  $x(t)$

$$x(t) = \sin\left(\frac{\pi t}{15}\right) + \gamma(t) \quad \gamma(t) \approx 500(0, \sigma_n^2)$$

$$\sigma_1^2 = 0.05 \quad \sigma_2^2 = 0.10 \quad \sigma_3^2 = 0.50$$

Figure 14. Plot of test models with the change of noise level

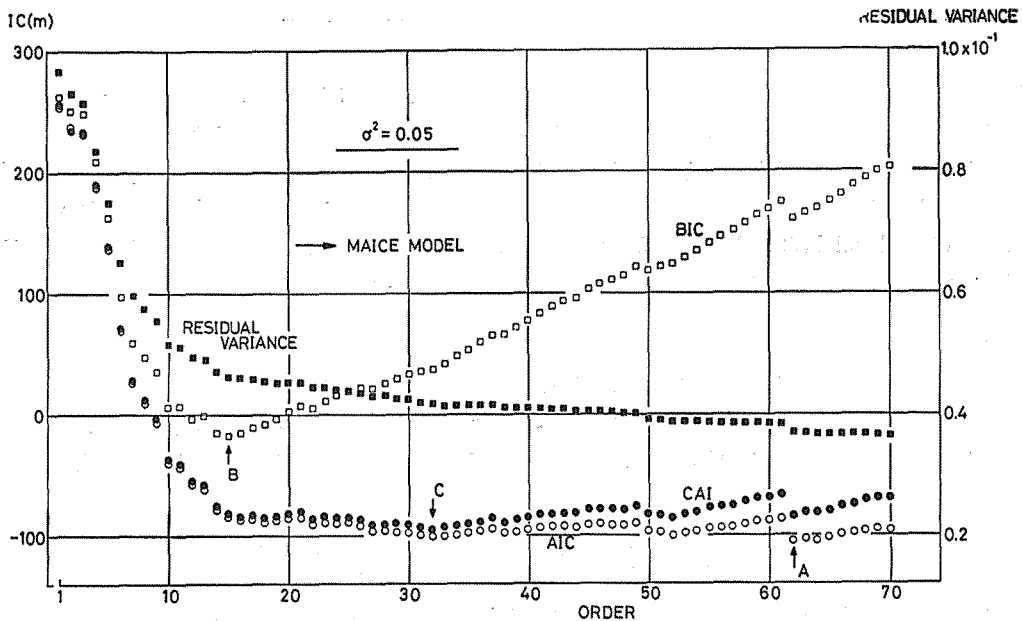


図-15  $\sigma_i^2=0.05$  の擬似系列に対する AR モデルのあてはめと残差分散, IC の次数に対するパス

■: 残差分散 ○: AIC □: BIC ●: CAI

Figure 15. Relationship between the IC estimation and residual variance with the change of order

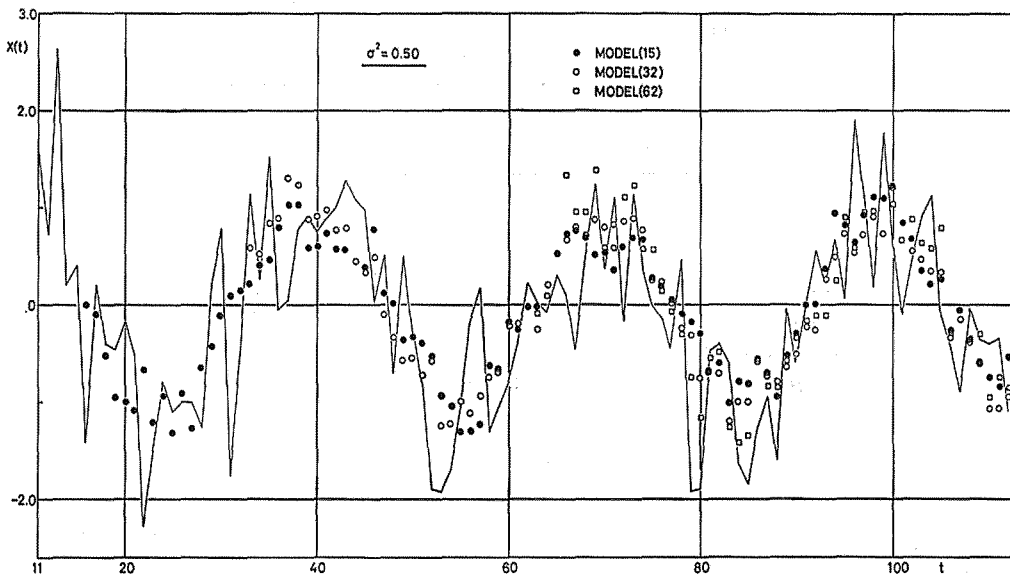


図-16 Model(15), Model(32), Model(62) による  $\sigma_3^2=0.50$  の擬似系列に対する予測  
—: 擬似系列  $x(t)$  ●: Model(15) ○: Model(32) □: Model(62)

Figure 16. Plot of predicted values for the test model ( $\sigma_3^2=0.50$ ) by the following AR models; Model(15), Model(32) and Model(62)

結果は、ノイズレベルに関係なくいずれのケースも **MAICE** モデルとして：**BIC=15 CAI=32 AIC=62** の **AR** モデルが選択された。真のモデルが正弦波であることを仮定しているの、各 **IC** はいずれも的確にモデルを選択していることになる。 $\sigma_1^2=0.05$  の場合の、残差分散と **IC** を次数との関係としてパスで比較したものが図-15である。残差分散の指数関数的減少傾向に対して、各 **IC** はこの残差分散を著しく減少させるモデルを敏感にキャッチしており、前述した各 **IC** の反応性と分解能の相異が、極めて明確に捉えられている。図-16は、各 **IC** から判定された **Model(15)**, **Model(32)**, **Model(62)** を用いて  $\sigma_2^2=0.50$  のノイズレベルの擬似系列  $x(t)$  を予測してプロットした一部を示しており、全体的に良く再現されていることがわかる。

## ②気象時系列の正規性の判定

雨量および気温時系列から、季節変動の型を取り去った系列に対して条件付正規性の同一性の判定を(24), (25)式に定義した **AIC** で検討する。各時系列に対するモデルを **Model( $\mu_1, \sigma_1^2$ )**, 正規乱数により得られるモデルを **Model( $\mu_2, \sigma_2^2$ )** とし、両者が異なる正規分布に従うモデルと仮定して

$$\text{Model}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y-\mu^2/2\sigma^2}$$

を考える。この **Model( $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ )** と **Model( $\mu, \mu, \sigma^2, \sigma^2$ )** に対して **AIC** を算定した結果が表-7であり、雨量時系列、気温時系列のいずれのモデルに対しても

$$\text{AIC}(\mu, \mu, \sigma^2, \sigma^2) < \text{AIC}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$$

が成り立ち、**Model( $\mu, \mu, \sigma^2, \sigma^2$ )** が採択される。従って、これらの系列は同一な正規分布に従うことが理解される。

表-7 季節変動パターンを除去した雨量・気温時系列に対する  
**AIC** による正規分布の同一性の判定

Table 7. Identification of time series eliminated the effect of seasonal change in case of monthly precipitation and air temperature with Gaussian process by **AIC** estimate

Model	A I C	
	Monthly total precipitation	Monthly mean of air temperature
$(\mu_1, \sigma_1^2)$	2131.71	745.54
$(\mu_2, \sigma_2^2)$	2125.16	748.21
$(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$	4256.87	1493.75
$(\mu, \mu, \sigma^2, \sigma^2)$	4254.87	1490.65

## V 結 語

気象データの時系列構造の把握と統計モデルによる予測という問題を、モデルの構成と情報量による評価という一貫した視点に立って議論してきた。気象データを対象としたのは、最近の気象変動様相が当演習林における造林地の雪害、林道の崩壊等の災害を少なからずもたらしており、将来的な森林管理計画に対して、予知・予測という形で何らかの基礎的な気象情報を利用できないかと考えたからである。

**IC** は、問題となる仮説を的確に表わす統計モデルを構成しさえすれば、これまで扱われたこ

とのない様な新しい問題にも自由に適用できる。 $\dot{I}C$ 最小化推定法の考え方は、既存のモデルに制約された検定でなく、既存のモデルの適切な利用と問題に応じた新しいモデル構成が重要であること、これにより、本来検定の目差すものが実現されるものであることを示唆している。

普段、何気なしに使っている統計的手法にも思わぬおとし穴があるかもしれないことを再認識したいと思う。

AIC に関する数式や記述法の多くは、赤池・坂元・田辺・北川等の論文から引用させて頂いたものであるが、浅薄な理解や曲解による部分があるとすれば、言うまでもなく著者等の浅学さによるものであり、率直な御批判を賜われれば幸いである。

本研究遂行にあたり深い御理解と有益な御指摘を頂いた鳥取大学神崎康一教授、北濱山本俊明助教授ならびに林業工学研究室の諸先生方に深謝したい。統計数理研究所の田辺國士先生には貴重な論文を提供していただいた。東南アジア研究センターの内田晴夫助手、大型計算機センターの高井孝之氏、かんがい排水研究室の大槻恭一氏には参考文献の紹介・収集について御協力を得た、併せて謝意を表したい。川那辺三郎教授をはじめとする当演習林の教職員の方々にも多くの励ましを頂台した、心から御礼申し上げたい。

なお、計算処理は京大大型計算機センターを利用したことを付記する。

## 引用文献

- 1) 田瀬則雄：水文時系列の解析とモデル化，古今書院，16—27，(1978)
- 2) 鈴木栄一：気象統計学，地人書館，(1973)
- 3) アタナシアス・パポリス：工学のための応用確率論，東海大学出版会，542—568，(1975)
- 4) 藤井光昭：時系列解析，コロナ社，1—211，(1977)
- 5) 北川源四郎・赤池弘次・大津皓平：局所定常自己回帰モデルによる適応制御系の実現，統計数理研究所集報，27，1，95—105，(1980)
- 6) 尾崎統：水文システムの統計モデル，数理科学，213，37—44，(1981)
- 7) Atkinson, A. C.: Posterior probabilities for choosing a regression model, *Biometrika*, 65, 1, 39-48, (1978)
- 8) Quenouille, M. H.: A large-sample test for the goodness of fit of autoregressive schemes, *J. Roy. Statist. Soc.*, 110, 123—129, (1947)
- 9) Anderson, T. W.: *The statistical Analysis of Time Series*, John Wiley & Sons, New York, (1971)
- 10) 日野幹雄：スペクトル解析，朝倉書店，25—227，(1977)
- 11) Akaike, H.: Power spectrum estimation through autoregressive model fitting, *Ann. Inst. Statist. Math.*, 21, 407-419, (1969)
- 12) ———: On the likelihood of a time series model, *The Statistician*, 27, 217-235, (1979)
- 13) 有本卓：確率・情報・エントロピー，森北出版，1—122，(1980)
- 14) 芝正己・古谷士郎・山本俊明・佐々木功：電算機を用いた数値地形解析による山岳地域の地形変動の計量化について(Ⅲ)，京大演報，55，194—221，(1983)
- 15) 芝正己・佐々木功：情報量規準 IC (Information Criterion) による統計モデルの評価と選択，95回日本林学会大会要旨，26，(1984)
- 16) 橋本明浩・本多正幸・井上隆勝・田栗正章：情報量規準による回帰モデルの選択，数理解析研究所講究録，378，52—60，(1979)
- 17) 田辺國士・田中輝雄：ベイズモデルによる曲線・曲面のあてはめ，月刊地球，5，3，179—186，(1983)
- 18) 中塚利直：時系列解析の数学的基礎，教育出版，22—228，(1978)
- 19) J. S. ベンダット・A. G. ピアソル：ランダムデータの統計的处理，培風館，9—379，(1977)
- 20) 坂元慶行・石黒真木夫・北川源四郎：情報統計学，情報統計学講座 A・5・4，共立出版，1—154，(1983)
- 21) Tong, H.: Determination of the order of a Markov chain by Akaike's Information Criterion, *J. Appl. Prob.* 12, 488—497, (1975)
- 22) Ozaki, T.: On the order determination of ARIMA models, *Appl. Statist.*, 26, 290-301 (1977)
- 23) Akaike, H.: Fitting autoregressive models for prediction, *Ann. Inst. Statist. Math.*, 21, 243-247,

- (1969)
- 24) 赤池弘次・中川東一郎: ダイナミックシステムの統計的解析と制御, サイエンス社, (1972)
  - 25) Jenkins, G. H.: Tests of hypotheses in the linear autoregressive model, *Biometrika*, **41**, 405-419, (1954)
  - 26) Haggan, V. and Ozaki, T.: Modeling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, *Biometrika*, **68**, 1, 189-196, (1981)
  - 27) Kitagawa, G.: On a search procedure for the optimal **AR-MA** order, *Ann. Inst. Statist. Math.*, **29B**, 2, 319-332, (1977)
  - 28) Akaike, H.: On the design of lag window for the estimation of spectra, *Ann. Inst. Statist. Math.*, **14**, 1, 1-21, (1962)
  - 29) 北川源四郎: 時系列解析—時間領域でのモデリングとその応用, *情報処理*, **21**, 11, 1174-1183, (1980)
  - 30) 柴田・田中: 積雪分布の推定, *研究時報*, **32**, 32, 51-57, (1980)
  - 31) 田瀬則雄: 月降水量の時系列構造と相互相関係数の関係, 筑波大学水理実験センター報告, **6**, 83-86, (1982)
  - 32) 朝倉正・久保光熙: 月降水の解析—降水量の長期予報 (第1報)—, *研究時報*, **13**, 9, 18-30, (1961)
  - 33) 田瀬則雄: 月降水量の相互相関係数に関する一考察, *ハイドロロジー*, **11**, 25-29, (1981)
  - 34) 柴田裕司: 数量化の手法による平均降水量の推定, *研究時報*, **21-5**, 44-50, (1969)
  - 35) 土木学会編: 土木工学における数値解析/計画手法編, サイエンス社, 56-81, (1975)
  - 36) 菅原正巳・赤池弘次: 水資源の変動様相に関する調査報告, 科学技術庁資源調査会報告, **34**, (1965)
  - 37) 菅原正巳・勝山ヨシ子: 気象資料の長期変動について, 防災科学技術センター研究報告, **1**, 45-70, (1968)
  - 38) 気象庁統計課: 任意地点の月平均気温の推定法, *気象庁技術報告*, **2**, (1960)
  - 39) 平野勝臣: 検定推定量の有意水準, *統計数理解析集報*, **27**, 1, 23-34, (1980)
  - 40) 稲垣宣生: 推定尤度関数と稜線推定量の一般化, *数理解析研究所講究録*, **380**, 1-15, (1980)
  - 41) 柴田里程: 密度関数の推定について, **380**, ———, 32-43, (1980)
  - 42) 渋谷政昭・柴田義真: 変化点の推定, **380**, ———, 83-100, (1980)

## Résumé

The **IC** (Information Criterion) is useful in selecting a model objectively when various models of different orders seem to fit equally.

In this paper the difficulty in deciding the order of an **AR** model (autoregressive model) in time series of the following meteorological data:

① daily mean of air temperature, ② annual record of maximum snow depth, ③ monthly total precipitation and ④ monthly mean of air temperature are discussed.

The results of the behaviour of autocorrelations are rather simple and any significant cycle cannot be found except the seasonal change in case of monthly precipitation and air temperature. To eliminate the effect of seasonal cycle, the time series is decomposed into two parts, the part of seasonal cycle and the remaining part with a slight persistence. Correlograms of the remaining part show a typical pattern of Markov process, therefore the advantage of this procedure lies in its use of the **AR** model which is one of the most versatile models of time series.

Even if without any other information for reducing the set of possible **AR** models, the behaviour of employed **IC** may suggest the assignment of some upper and lower limits of **AR** models.

It is remarkable that these steps overcome many of the difficulties of the order determination of **AR** models and provide a basis for the development of identification of higher order **AR** models.